

CLAIMS

What is claimed is:

- 5 1. A method in a data communications device for directing a request to process data, comprising the steps of:
- receiving the request from a client;
- selecting one of a plurality of resources to process the request from a plurality of resources based on generating an estimated response usage for each resource of the
- 10 plurality of resources that reflects a potential usage if responding to the request; and
- forwarding the request to the one of the plurality of resources selected in the step of selecting.
2. The method of claim 1, wherein the step of selecting the one of the plurality of
- 15 resources comprises:
- generating an estimated available usage for each resource based on the request and usage information received from a usage meter for each resource; and
- selecting the one of the plurality of resources based on a highest estimated available usage for each resource.

20

3. A data communications device for directing a request to process data, comprising:
a memory that stores a cost modeler application;
an interconnection mechanism; and
a processor coupled to the memory by the interconnection mechanism, wherein
- 5 the processor operates in accordance with instructions of the cost modeler application
stored in the memory to direct the request, the instructions of the cost modeler application
configuring the processor to:
- receive the request from a client;
 - select one of a plurality of resources to process the request from a plurality

10 of resources based on generating an estimated response usage for each resource of
the plurality of resources that reflects a potential usage if responding to the
request; and

 - forward the request to the one of the plurality of resources selected to
process the request.
- 15
4. The data communications device of claim 3, wherein the cost modeler application
comprises further instructions that configure the processor to:
- generate an estimated available usage for each resource based on the request and
usage information received from a usage meter for each resource; and

20 select the one of the plurality of resources based on a highest estimated available
usage for each resource.

5. A data communications device for directing a request to process data, comprising:
- means for receiving the request from a client;
 - means for selecting one of a plurality of resources to process the request from a plurality of resources based on generating an estimated response usage for each resource of the plurality of resources that reflects a potential usage if responding to the request;
 - and
 - means for forwarding the request to the one of the plurality of resources selected in the step of selecting.
6. A computer program product that includes a computer readable medium having instructions stored thereon for directing a request to process data, such that the instructions, when carried out by a data communications device, cause the data communications device to perform the steps of:
- receiving the request from a client;
 - selecting one of a plurality of resources to process the request from a plurality of resources based on generating an estimated response usage for each resource of the plurality of resources that reflects a potential usage if responding to the request; and
 - forwarding the request to the one of the plurality of resources selected in the step of selecting.

7. A method for selecting a resource from a plurality of resources to process a request from a client, comprising the steps of:

generating a first cost increase for a first resource of the plurality of resources if the first resource responds to the request and a second cost increase for a second resource of the plurality of resources if the second resource responds to the request;

comparing the first cost increase and the second cost increase to determine one of the first and second cost increases that has a lower cost increment; and

selecting one of the first resource and the second resource to respond to the request from the client based on the lower cost increment.

8. The method of 7, wherein the step of comparing the first cost increase and the second cost increase comprises determining that the first cost increase exceeds a preset first cost level and the second cost increase does not exceed a second preset cost level.

9. A data communications device for selecting a resource from a plurality of resources to process a request from a client, comprising:

a memory that stores a cost modeler application;

an interconnection mechanism; and

5 a processor coupled to the memory by the interconnection mechanism, wherein the processor operates in accordance with instructions of the cost modeler application stored in the memory to select the resource, the instructions of the cost modeler application configuring the processor to:

generate a first cost increase for a first resource of the plurality of

resources if the first resource responds to the request and a second cost increase for a second resource of the plurality of resources if the second resource responds to the request;

compare the first cost increase and the second cost increase to determine one of the first and second cost increases that has a lower cost increment; and

15 select one of the first resource and the second resource to respond to the
request from the client based on the lower cost increment.

10. The data communications device of claim 9, wherein the cost modeler application comprises further instructions that configure the processor to determine that the first cost increase exceeds a preset first cost level and the second cost increase does not exceed a second preset cost level.

11. A data communications device for selecting a resource from a plurality of resources to process a request from a client, the data communications device comprising:

means for generating a first cost increase for a first resource of the plurality of resources if the first resource responds to the request and a second cost increase for a
5 second resource of the plurality of resources if the second resource responds to the request;

means for comparing the first cost increase and the second cost increase to determine one of the first and second cost increases that has a lower cost increment; and

means for selecting one of the first resource and the second resource to respond to
10 the request from the client based on the lower cost increment.

12. A computer program product that includes a computer readable medium having instructions stored thereon for selecting a resource from a plurality of resources to process a request from a client, such that the instructions, when carried out by a data
15 communications device, cause the data communications device to perform the steps of:

generating a first cost increase for a first resource of the plurality of resources if the first resource responds to the request and a second cost increase for a second resource of the plurality of resources if the second resource responds to the request;

comparing the first cost increase and the second cost increase to determine one of
20 the first and second cost increases that has a lower cost increment; and

selecting one of the first resource and the second resource to respond to the request from the client based on the lower cost increment.

13. A method for selecting a resource from a plurality of resources to process a request from a client, comprising the steps of:

generating a usage metric for each resource of the plurality of resources;

5 generating an economic metric for each resource based on the usage metric for each resource and the request; and

choosing one of the plurality of the resources to respond to the request for data based on a comparison of the economic metric for each resource.

14. The method of claim 13, wherein the step of generating the usage metric
10 comprises the steps of:

receiving usage information from a usage meter for each resource that measures the amount of usage of each resource over time;

generating an estimated base usage of each resource based on the usage information; and

15 generating a peak usage metric for each resource that represents a measurement of the highest level of usage attained for each resource over a current billing period based on the usage information.

15. The method of claim 13, wherein the step of generating the economic metric
20 comprises the steps of:

generating an estimated request usage for each resource that reflects an estimate of the projected usage of each resource over a current time based on the request for data; and

25 generating an estimated available usage for each resource based on the usage metric for each resource and the estimated request usage for each resource.

16. The method of claim 15, wherein:

the step of generating the estimated available usage comprises generating the estimated available usage based on the estimated request usage for each resource, an estimated base usage for each resource based on usage information received for each resource, and a peak usage metric for each resource that represents a measurement of the highest level of usage attained for each resource over a current billing period based on the usage information.

17. The method of claim 16, wherein the step of generating the estimated available usage comprises the steps of:

identifying a subset of the plurality of resources wherein the estimated request usage is higher than a predetermined usage limit defined relative to the peak usage metric for each resource; and

assigning a predetermined marginal cost to an incremental usage of each resource in the subset.

18. The method of claim 13, wherein the step of generating the economic metric comprises generating a bandwidth metric that represents the bandwidth requirements for each resource based on (i) the request and (ii) the network address of a client providing the request and the network address of each resource.

19. The method of claim 13, wherein the step of generating the economic metric comprises generating the economic metric based on a category of the request that indicates an estimated request usage for each resource.

20. The method of claim 13, wherein the step of choosing one of the plurality of resources comprises the step of:

comparing an estimated available usage for each resource generated based on the request and the usage metric for each resource.

5

21. A data communication device for selecting a resource from a plurality of resources to process a request from a client, the data communication device comprising:

a memory that stores a cost modeler application;

an interconnection mechanism; and

10 a processor coupled to the memory by the interconnection mechanism, wherein the processor operates in accordance with instructions of the cost modeler application stored in the memory to select the resource, the instructions of the cost modeler application configuring the processor to:

generate a usage metric for each resource of the plurality of resources;

15 generate an economic metric for each resource based on the usage metric for each resource and the request; and

choose one of the plurality of the resources to respond to the request for data based on a comparison of the economic metric for each resource.

20 22. The data communications device of claim 21, wherein the cost modeler application comprises further instructions that configure the processor to:

receive usage information from a usage meter for each resource that measures the amount of usage of each resource over time;

25 generate an estimated base usage of each resource based on the usage information; and

generate a peak usage metric for each resource that represents a measurement of the highest level of usage attained for each resource over a current billing period based on the usage information.

23. The data communications device of claim 21, wherein the cost modeler application comprises further instructions that configure the processor to:

generate an estimated request usage for each resource that reflects an estimate of the projected usage of each resource over a current time based on the request for data;

5 and

generate an estimated available usage for each resource based on the usage metric for each resource and the estimated request usage for each resource.

24. The data communications device of claim 23, wherein the cost modeler application comprises further instructions that configure the processor to generate the estimated available usage based on the estimated request usage for each resource, an estimated base usage for each resource based on usage information received for each resource, and a peak usage metric for each resource that represents a measurement of the highest level of usage attained for each resource over a current billing period based on the usage information.

15

25. The data communications device of claim 24, wherein the cost modeler application comprises further instructions that configure the processor to:

identify a subset of the plurality of resources wherein the estimated request usage is higher than a predetermined usage limit defined relative to the peak usage metric for each resource; and

20

assign a predetermined marginal cost to an incremental usage of each resource in the subset.

25

26. The data communications device of claim 21, wherein the cost modeler application comprises further instructions that configure the processor to generate a bandwidth metric that represents the bandwidth requirements for each resource based on (i) the request and (ii) the network address of a client providing the request and the network address of each resource.

27. The data communications device of claim 21, wherein the cost modeler application comprises further instructions that configure the processor to generate the economic metric based on a category of the request that indicates an estimated request usage for each resource.

28. The data communications device of claim 21, wherein the cost modeler application comprises further instructions that configure the processor to compare an estimated available usage for each resource generated based on the request and the usage metric for each resource.

29. A data communications device for selecting a resource from a plurality of resources to process a request from a client, the data communications device comprising:
means for generating a usage metric for each resource of the plurality of resources;
means for generating an economic metric for each resource based on the usage metric for each resource and the request; and
means for choosing one of the plurality of the resources to respond to the request for data based on a comparison of the economic metric for each resource.

30. A computer program product that includes a computer readable medium having instructions stored thereon for selecting a resource from a plurality of resources to process a request from a client, such that the instructions, when carried out by a data communications device, cause the data communications device to perform the steps of:
- 5 generating a usage metric for each resource of the plurality of resources;
 generating an economic metric for each resource based on the usage metric for each resource and the request; and
 choosing one of the plurality of the resources to respond to the request for data based on a comparison of the economic metric for each resource.